DOCUMENT RESUME

ED 278 550 SE 047 714

AUTHOR	Shymansky, James A.; And Others
TITLE	A Reassessment of the Effects of 60's Science Curricula on Student Performance: Final Report.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	[87]
GRANT	MDR-8550470
NOTE	58p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; Instructional Improvement; *Meta Analysis; *Research Methodology; *Science Curriculum; Science Education; *Science Programs; *Statistical Analysis; Student Attitudes
IDENTIFIERS	*Science Education Research

ABSTRACT

A meta-analysis of studies of the impact of new science curricula on the performance of students was initially conducted in 1983 as part of a coordinated program of meta-analysis of research on science education. This paper reports the results of a re-synthesis of the research dealing with student performance in the new science curriculum using refined statistical procedures developed by Hedges and Hedges and Olkin. A comparison of procedures in the previous meta-analysis may have led to misleading results. Descriptions are provided of this study's: (1) problem formulation; (2) data collection and evaluation techniques; and (3) data analysis. Results suggest that the science programs developed during the sixties and seventies were generally effective in improving student performance on cognitive measures and in raising attitudes about science. Data are summarized in 27 tables. (ML)

A REASSESSMENT OF THE EFFECTS OF 60'S SCIENCE CURRICULA ON STUDENT PERFORMANCE: FINAL REPORT[1]

James A. Shymansky
Science Education Center
The University of Iowa

Larry V. Hedges
Department of Education
The University of Chicago

George Woodworth
Department of Statistics
The University of Iowa

[1]This report is based upon research supported by the National Science Foundation under Grant No. MDR-8550470. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation.

A REASSESSMENT OF THE EFFECTS OF 60'S SCIENCE CURRICULA ON STUDENT PERFORMANCE: FINAL REPORT

James A. Shymansky
Science Education Center
The University of Iowa

Larry V. Hedges
Department of Education
The University of Chicago

George Woodworth
Department of Statistics
The University of Iowa

INTRODUCTION

The advent of new elementary and secondary science curricula in the 1960s and 1970s led to a large number of research studies designed to determine the impact of these curricula on students. An initial search of the literature revealed more than 300 research studies and discussion papers on the topic. The number and variety of these studies is so great as to defy simple narrative summary. To understand this body of research as a whole, an alternative to traditional narrative review is needed. One alternative is the use of a quantitative research synthesis method called meta-analysis (Glass, 1976).

Meta-analysis usually involves summarizing the results of each study by an estimate of effect magnitude and then combining these estimates across studies (Glass, McGaw , and Smith, 1981). Meta-analysis permits the research reviewer to utilize the results of large numbers of research studies while preserving information on the ways in which studies differ via indicators of study, population, and treatment characteristics. These indicators of between study differences can be used to determine whether the results of studies (the effect magnitudes) are systematically related to differences between studies.

A meta-analysis of studies of the effects of new science curricula on the performance of students was conducted by Shymansky, Kyle, and Alport (1983) as part of a coordinated program of meta-analyses of research on science education (Anderson, Kahl, Glass, and Smith, 1983). The results of the Shymansky, et al., study suggested many interesting conclusions about the effects of the new curricula. These included generally positive effects on cognitive and affective performance, but some variations in the effects were found for various subgroups of students.

The methods used in the Anderson project were essentially those proposed by Glass (1976) and elaborated by Glass, McGaw, and Smith (1981). Although these methods represented the state of the art at the time, improvements in conceptualization and statistical theory for meta-analysis suggest that the results generated in the Shymansky, et al., study and other earlier syntheses may have been in error, improperly estimating the effect magnitude, precision, and possibly the effect direction (See Hedges, 1981, 1982a, 1982b, 1982c; Hedges and Olkin, 1985; Kraemer, 1983; Rosenthal and Rubin, 1982). Considering the significance of the research on new science curricula in decisions of future curriculum design (NSF Guidelines, 1986), it seems only prudent to reanalyze the research with the more refined statistical methods now available.

This paper reports the results of a re-synthesis of the research dealing with student performance in the new science curricula using the refined statistical procedures proposed by Hedges (1981, 1982a, 1982b, 1982c) and by Hedges and Olkin (1985). A comparison of procedures used and results obtained in the original Shymansky, et al., study and the present re-analysis is provided.

BACKGROUND

Our re-analysis was prompted by the development in the last 5 years of sophisticated statistical methods for meta-analysis. In this section we describe why the use of conventional statistical procedures in the previous meta-analysis may have lead to misleading results.

Problems with conventional statistical methods in meta-analysis. There are two types of problems with the use of conventional statistical procedures in meta-analysis. The first problem is essentially conceptual. Conventional meta-analyses give equal weight to all studies regardless of the precision with which they estimate the treatment effect. Moreover, these analyses do not permit all of the statistical tests that are of interest and that are necessary to understand completely the variation in a set of effect size data. The second problem is statistical. Conventional meta-analyses involve very serious violations of the assumptions of statistical procedures such as t-tests, analysis of variance, and correlational analysis. These violations are likely to have a profound effect on statistical significance levels casting doubt on the results of the meta-analysis (see Hedges and Olkin, 1985).

Conceptual problems with the application of conventional statistical methods in meta-analysis. Conventional statistical methods implicitly assume that each observation in a data set has unknown, but equal, precision. That is, the variance of the sampling error associated with each observation is assumed to be the same for all cases. In conventional meta-analysis procedures, all effect sizes are treated as though they have equal precision. It has been shown analytically, however, that the precision (error variance) of an effect size estimate is inversely proportional to the sample size of the study on which it is based (Hedges, 1981). When these sample sizes differ across studies (as they do in

almost every case), the effect size estimates will also differ in precision. Consequently, meta-analytic methods that do not recognize differences in precision are likely to be misleading for purely conceptual reasons.

Reviewers of research usually want to test both whether the main effect of treatment (the average treatment effect size) is different from zero and whether the treatment effect is consistent across studies. Although it is possible to construct a conventional test that the average effect size is zero, this test is problematic for reasons that will be discussed shortly. Moreover, it is impossible to test directly the consistency of effect sizes across studies in the conventional meta-analysis. The conventional analysis for testing systematic variation among k effect sizes would need k-1 degrees of freedom for systematic variation among effect sizes and one degree of freedom for the grand mean, the result being there are no degrees of freedom remaining for estimation of the error or nonsystematic variation. In the conventional framework, it is therefore impossible to construct an omnibus test to determine whether the systematic variation in k effect sizes is larger than the nonsystematic variation exhibited by those effect sizes.

Using conventional methods it is possible to test that the average effect size is zero (or that the differences among the average effects sizes of two or more groups of studies are zero) as long as at least one group contains two or more effect sizes. The multiple effect sizes within the group(s) serve as replicates from which an estimate of nonsystematic variance is obtained. Then the test is constructed by comparing the "systematic" variance of effect sizes around zero (or among group mean effect sizes) to the "nonsystematic" variance of effect sizes within groups.

However, such a test is conceptually and statistically perilous. The investigator does not know if the effect sizes exhibit only nonsystematic variability within groups. If the investigator chooses the wrong groups, considerable systematic variance may be pooled into the estimate of error variance. Moreover, conventional wisdom suggests that there are many reasons to expect systematic variation between study results due to differences in study design, treatment implementation, and subject samples (see e.g., Presby, 1978). Including systematic variance in the estimates of error terms decreases the sensitivity of the statistical test for systematic variation (Madow, 1948).

The conceptual problem in conventional meta-analysis is that the amount of systematic variation among observed effect sizes is unknown. Moreover, even if the test for specific contrasts among effect sizes were not problematic, such tests could not substitute for a direct omnibus test of the consistency of study results. Omnibus tests for the consistency of study results test the hypothesis that all possible contrasts are zero, including constrasts not anticipated by the investigator.

Statistical problems with application of conventional statistical methods in meta-analysis. The analysis of effect sizes or correlation coefficients by using conventional statistical methods is also problematic for purely statistical reasons. Conventional procedures (t-tests, analysis of variance, multiple regression analysis) rely on parametric assumptions about the data. All of these procedures require that the nonsystematic variance associated with every observation be the same (the so-called homoscedasticity assumption). That is, if each observation is composed of a systematic part and an error part, then the errors for all observations must be equally variable. In analysis of variance this assumption is checked by examining the within-cell variances to see if they are similar in

value. In regression analysis this assumption is checked by determining whether the residual variance about the regression line is reasonably constant for all values of the predictor variable.

When the observations are estimates of effect magnitudes (either effect sizes or correlation coefficients), statistical theory provides exact values for the nonsystematic variance of each observation. Hedges (1981) has shown that the nonsystematic variance of an effect size estimate is inversely proportional to the sample size on which the estimate is based. Thus, as sample sizes of studies vary over a wide range, so will error variances. Since it is not unusual for within-study sample sizes in meta-analyses to differ by a factor of 50, substantially heterogeneous error variances are possible.

The effect of heterogeneity of variance on analysis of variance F tests have been studied extensively (see e.g., Glass, Peckham, and Sanders, 1972). Furthermore, heterogeneous variances have only small effects on the validity of F tests in most applications of analysis of variance. However, the situation in research synthesis is quite different from that in which robustness of F test is usually studied. Studies of the effects of heterogeneity of variance in ANOVA usually give a different variance to one or more groups in the design. Thus every observation in the same group has the same variance, and there are at most two to three different variances in the entire experiment.

In the case of research synthesis the heterogeneity is usually more pronounced. Every observation (study) may have a different variance. Moreover, the range of variances studied in connection with the robustness of F test is usually rather limited, often less than 5 to 1. The studies that examine the effects of very wide ranges of variances and groups of unequal size find that the F test is not necessarily robust to substantial heterogenity of variance. For

example, Glass, Peckham, and Sanders (1972) note that when the ratio of variances is 5 to 1 and the sample sizes are unequal, the actual significance level of the F test can be six times as large as the nominal significance level, say 0.30 instead of 0.05.

Thus, the violation of the homogeneity of variance assumption of analysis of variance or regression analysis is severe in research synthesis. Though this type of violation has not been studied extensively, it is mathematically equivalent to the violation of assumptions that occurs when the error terms are correlated (that is, when observations are not independent) -- a violation of assumptions against which the F test is demonstrably not robust (See Box, 1954; Scheffe, 1959). Thus, there is little reason to believe that the usual robustness of the F tests will somehow prevail. The statistical problem of violation of the assumptions of conventional statistical procedures and the potential problem of bias due to pooling of systematic variation into estimates of error variance raise severe questions about the validity of conventional statistical procedures in meta-analysis. There is no rigorously defensible argument for the general use of conventional t-tests, analysis of variance, or regression analysis to analyze effects sizes and correlations in most meta-analyses. These procedures can produce spurious results and should no longer be used now that an extensive set of valid statistical procedures is available.

PROCEDURES

Methodological standards in original research help ensure the validity of the research. The same standards should apply to research syntheses. As in original research, a meta-analysis can be conceptualized as having five stages: problem formulation, data collection, data evaluation, data analysis and interpretation, and

presentation of results. Since this study is, in a sense, a replication of the earlier study by Shymansky, et al. (1983), certain decisions regarding problem formulation, data collection, and presentation of results were determined by that study. The parallel construction of this study facilitates a direct comparison of results obtained using conventional meta-analytic methods and the more refined procedures.

Problem formulation

The Shymansky, et. al (1983) meta-analysis was designed to determine the impact of new science curricula on student performance. This impact was analyzed using the effect sizes calculated from student performance on 18 types of criterion measures.

In the original study, new science curricula were defined as those programs which:

(a) were developed after 1955 (with either public or private funds),

(b) emphasized the nature, structure, and processes of science,

(c) integrated laboratory activities into the core of the instruction, and

(d) emphasized higher cognitive skills and an appreciation of science.

Traditional curricula were defined as those programs which:

(a) were developed or patterned after a program developed prior to 1955,

(b) emphasized knowledge of scientific facts, laws, theories and applications, and

(c) used laboratory activities as verification exercises or as lesson supplements only.

In the original study problems were often encountered in applying the criteria for new and traditional curricula when the studies mentioned that programs were modified but did not describe the exact nature of the modifications (Shymansky,

et al., 1983, p. 389). This problem was evident in the re-analysis when each of the 105 studies included in the original analysis was examined more closely. For example, several studies compared students who had been exposed to a new science program with students who had been exposed to no science program. Although there might be some who could argue that "no science" is traditional in many schools (especially elementary school), these conditions clearly do not meet the intent of the meta-analysis. In the re-analysis the data from these studies were eliminated from the analyses as they did not meet the intent of the synthesis (i.e., comparing new and old science curricula).

In the original meta-analysis the 18 student performance measures were grouped into six "criterion clusters": science achievement, student perceptions, process skills, problem solving, related skills (reading, math computation, writing), and other performance areas (involving mostly studies of Piagetian task performance). These criterion clusters were used as the dependent variables in this meta-analysis. Separate statistical analyses were used to examine the effects of new science curricula on each of these six criterion clusters.

Studies were also differentiated on the basis of the particular new curriculum studied, the implementation strategies utilized, study design characteristics, and data on student background. Characteristics of each study which could potentially influence the estimated effect of the new science curriculum were recorded. These variables include the design of the study (randomized or other), type of criterion measure (ad hoc or standardized), the nature of the criterion measure (cognitive, affective, etc.), identity of the new science curriculum studied (ESS, SCIS, PSSC, etc.), length of the trial of the new curriculum, degree of teacher inservice or preservice, school size and type, and students' average age, IQ, socioeconomic status, and class male/female ratio.

Data collection and evaluation

The 105 studies included in the original Shymansky, et al., study constituted the sample for the re-analysis. Copies of the original manuscripts were retrieved and re-coded by two independent coders. Where disagreement or problems of study coding occurred, the full research team met to discuss and decide on the final disposition of the study. In the process of re-coding the studies, we found that statistical reporting practices were far from uniform and in some cases fell below minimal standards. If a study fails to report means, standard deviations, and sample sizes for each treatment group (pupils taught under a new curriculum) and control group (pupils taught under a traditional curriculum), a weighted effect size estimate cannot be calculated. Although some studies which failed to provide this information were useable in the original study because it was possible to construct an effect size estimate by algebraic transformations, many studies could not be used with the refined procedures. Among the other specific problems encountered in reviewing the original sample of 105 were studies which: (a) matched inappropriate comparison groups, (b) reported only results that used the class as the unit of analysis, (c) reported only analyses based on gain scores, or (d) reported conflicting, inconsistent data. This process resulted in the elimination of 24 studies for the re-meta-analysis or a final sample size of 81 studies.

Data analysis

How was the impact of the new curricula quantified? Imagine that a traditional and new curriculum are taught to large groups of pupils drawn from the same population. At the end of instruction, each pupil is assessed on some criterion measure. The impact of the new curriculum in terms of an effect size estimate is the difference between the mean criterion scores for the new and

traditional curricula expressed in standard deviation units. We used the standard deviation under the traditional curriculum as the baseline, although other choices are possible (See Hedges and Olkin, 1985). For example, an effect size of 1.0 indicates that a criterion score which would be at the 50th percentile under the new curriculum would have been at the 84th percentile under the old curriculum; the new curriculum raises "C" pupils almost to "A" status. Clearly, experience and common sense would suggest that effect sizes greater than 1.0 ought to be viewed with skepticism.

Effect Size Estimates In practice, effect sizes are estimated from comparatively small samples from the relevant population. When means, standard deviations, and sample sizes were reported, we estimated effect sizes using the formulas:

$$(1) \quad g = (X_T - X_C)/S_c$$

and

$$(2) \quad d = Jg$$

where $J = 1 - 3/(4df - 5)$ is a bias correction factor, X_T and X_C are treatment and control sample means, S_c is the control sample standard deviation, n_T and n_C are treatment and control sample sizes, $df = n_C - 1$, and g and d are raw- and bias-corrected effect size estimates.

When means and standard deviations were not reported but an unadjusted t or a single factor ANOVA was reported, the formula (2) was used with

$$(3) \quad g = \pm [n_T n_C F_{TRT}/(n_T + n_C)]^{1/2}$$

and

(4) $df = n_T + n_C - 2$

where the sign is positive if treatment is better than control and negative if treatment is worse than control and F_{TRT} is the F test statistic or the square of the t statistic for the treatment effect. A few studies reported F-statistics but did not say if the new curriculum was better or worse than the traditional. These studies were omitted because their results were ambiguous.

When only the covariate adjusted means and standard deviations, a covariate-adjusted t statistic, or a multifactor ANOVA or ANCOVA table were reported, then (2), (3) and (4) were used with F_{TRT} defined by

(5) $F_{TOT} = (df_{TOT} - 1)SS_{TOT}/(SS_{TOT} - SS_{TRT})$

where SS_{TOT} is the total sum of squares about the mean, SS_{TRT} is the treatment sum of squares, and df_{TOT} is the total degrees of freedom (see e.g., Glass, McGaw, and Smith, 1981).

Effect size estimates calculated from covariate adjusted F statistics or adjusted sums of squares implicitly involve a covariate adjusted standard deviation. This tends to make such effect sizes larger than they would have been if they had been computed using an unadjusted (raw posttest score) standard deviation (see Glass, McGaw, and Smith, 1981). To correct for this artifact in the very small number of cases in which effect size could only be computed from covariate adjusted statistics, such effect size estimates and their calculated standard errors were multiplied by 0.70. This correction was derived

by assuming a correlation of about 0.71 between the covariate and the posttest.[1]

We also estimated the sampling standard error of each effect size; that is, the uncertainty in the effect size due to the comparative smallness of the sample sizes actually used in a study. Sampling standard error is the standard deviation of the estimated effect size around the true effect size of the population of pupils from which the study population was selected. In other words, it measures the sampling variation of the estimated effect size but does not reflect non-sampling variations which would occur if the study had used a different population of pupils or different teachers. The standard error of the estimated effect size, d, is given by the formula

$$s^2 = \frac{J^2 df(n_T + n_C)}{(df-2)n_T n_C} + \frac{d^2 (J^2 df)}{df-2} - d^2 \tag{6}$$

where $df = n_C - 1$ is the degrees of freedom for the control group. When total sample size was reported but treatment and control sample sizes were not reported separately, we assumed they were equal.

Statistical treatment of studies yielding more than one effect size. A substantial number of studies provided two or more effect size estimates. For example, a study may have reported statistics for third and fourth grade girls and boys on an achievement criterion measure and an attitude criterion measure, thus providing eight effect size estimates. In this case there are four subgroups (third and fourth grade girls and boys) and two subscales (achievement and attitude). Effect sizes from different subgroups are statistically independent

---

[1] The correction factor for the standard error is $(1-R^2)^{1/2}$, where R is the correlation between the posttest and the covariate.

while effect sizes on different subscales measured on the same subgroup are statistically correlated (although the correlation in some cases may be small).

Suppose that in a particular study subscales A, B, and C are to be aggregated. The individual effect size estimates and their standard errors are d_A, s_A; d_B, s_B; and d_C, s_C. The pooled effect size estimate is the average,

$$(7)\quad d = (d_A + d_B + d_C)/3.$$

It can be proved that regardless of the correlations among the subscales, the standard error of this pooled estimate is less than or equal to the average of their three standard errors,

$$(8)\quad s_d \leq (s_A + s_B + s_C)/3;$$

consequently, the average standard error is a conservative estimate of the standard error of the pooled effect size. This conservative standard error was used in all analyses where results were pooled across subscales.

Effect sizes derived from scores on criterion measures administered to students exposed to the traditional and new science curricula constituted the dependent variables in the meta-analysis. Criterion measures were grouped into one of five criterion clusters (achievement, perceptions, process skills, analytic skills, and other performance areas) as previously mentioned. However, in many cases two different criterion measures within one study fell into the same criterion cluster. For example, a standardized general science achievement test and a standardized chemistry test administered to the same students would both fall in the achievement cluster. In cases like this it was necessary to combine

these correlated effect sizes into one pooled effect size, since we had no basis for selecting one criterion to represent the cluster (as recommended by Hedges and Olkin, 1985, Chapter 10, Section G). We computed the unweighted average of the correlated effect size estimates where they occurred and made a conservative estimate of the standard error of their average as described previously. After such pooling, our data consisted of at most one effect size estimate for each criterion cluster for each independent student subgroup for each study.

When overall student performance comparisons were made (across criterion clusters), further pooling was done to produce at most one effect size for each independent subgroup of students for each study. This pooling of effects sizes to the level of independent student subgroups for each study is an important difference between this and the previous analysis of these data.

We grouped effect sizes into categories having similar study characteristics; for example, the same curriculum, student male/female ratio, or average IQ. Within each category we computed a weighted average effect size, in which each effect size estimate was weighted by its precision (the inverse of the squared sampling standard error as recommended by Hedges and Olking, 1985). In addition, we computed the total standard error, incorporating sampling and non-sampling components of variation.

Analysis of Heterogeneity We used single and multiple factor weighted ANOVA to calculate Hedges and Olkin's "Q" statistics (1985, Formula 18, Chapter 7, Section C.3). For example, two lines of data taken from the curriculum by criterion cluster analysis (Table 1) indicate there were 18 achievement effect size estimates for BSCS-Yellow from independent subgroups of students. The weighted average effect size was 0.47 (weighted by the inverse of the squared

sampling standard error). The sampling standard error of the average effect size was 0.02. However, the heterogeneity statistic (Q) was highly significant, indicating considerable non-sampling variation (heterogeneity) among the studies generating these 18 effect sizes.

In contrast, the two achievement effect size estimates for BSCS-Advanced produced a weighted average effect size of 0.11 with a sampling standard error of 0.12. However, the Q statistic was only 0.07 with 1 degree of freedom (which is non-significant at the 0.05 level), indicating homogeneity among the study results.

TABLE 1

PARTIAL OUTPUT OF CURRICULUM BY ACHIEVEMENT ANALYSIS

CURRICULUM	N[a]	MEAN ES[b]	SSE[c]	TSE[d]
BSCS-Y	18	0.47	0.02	0.11
BSCS-A	2	0.11	0.12	0.12

a. Number of statistically independent effect size estimates
b. Weighted mean effect size
c. Standard error of weighted mean effect size based on sampling error only
d. Standard error of weighted mean effect size based on both sampling and nonsampling variation

The goal of our analysis was to examine systematic differences between studies (e.g., differences in curricula, teacher preparation, student gender, etc.) which might account for the non-sampling variability. Even so, many of the aggregations produced heterogeneous results.

<u>Quantifying Nonsampling Variability</u> For descriptive purposes it is convenient to model what we have called "nonsampling" effect size variability as random. In a <u>random effects</u> model, it is supposed that the studies which produced the observed effect sizes are a random sample from a conceptual "hyperpopluation" of studies which might have been performed (see Hedges and Olkin, 1985). The <u>component of variance</u> due to random sampling of studies can then be estimated from a sample of studies from this hyperpopulation.

For example, the average effect size for BSCS-Yellow in 18 studies in Table 1 had a sampling standard error of 0.02. This figure encompasses variations due to sampling of students within the 18 study populations but fails to reflect variations among the populations themselves. When the latter component of variance is included, the total standard error becomes 0.11.

Clearly, the random effects model is reasonable only if the data have been broken to the point that no further non-sampling variation can be associated with observed characteristics; however, we feel that the combined estimate of standard error is useful in that it inspires caution in assessing the significance of the difference between two effect sizes. For example, referring to Table 1 again, the average effect sizes and sampling standard errors for BSCS-Yellow and BSCS-Advanced were 0.47, 0.02 and 0.11, 0.12, respectively. The sampling standard error of the difference between BSCS-Yellow and BSCS-Advanced is the square root of the sum of the squares of their individual sampling errors, i.e., 0.12. One might erroneously conclude that the two curricula produced effects differing by 3.00 standard errors ((0.47-0.11)/0.12). However, when the total standard errors (non-sampling and sampling variation) of 0.11 and 0.12 (for BSCS-Yellow and BSCS-Advanced, respectively) are taken into consideration, they differ by only 2.25 standard errors ((0.47-0.11)/0.16). Thus, the difference

between curricula is much less statistically significant when nonsampling standard errors are taken into account.

Diagnostic Plots We employed diagnostic plots and Q statistics (Hedges and Olkin, 1985, Chapter 12, Sections B.1 and B.3) to detect individual studies which deviated to an unusal extent from the group means. This proved useful in identifying miscoded studies (which were recoded and used), or methodologically defective studies (which were not included in this study).

RESULTS AND DISCUSSION

The results presented in this section parallel those in the earlier study by Shymansky, et al. (1983). But because results from 24 of the studies included in the 1983 synthesis could not be used (for reasons explained earlier), the 1983 analyses had to be repeated on the surviving 81 studies to illustrate the effect of the refined statistical methods on the outcome. Tables 2-13, therefore, list the results of the unweighted procedures on the original 105 studies, the results of the unweighted procedures on the "surviving" 81 studies, and the results of the weighted procedures on the 81 studies.

We analyzed the variation of effect sizes across studies to address the following questions: How much do the study characteristics (between study independent variables) influence the effects of new curricula? Do these influences account for most of the non-sampling variation? How much non-sampling variation cannot be accounted for? The methods we used involve weighted estimates of average effect sizes within categories of studies, analyses of heterogeneity, diagnostic plots, and statistics to locate deviant studies.

In general we found significant heterogeneity in many cases which shows that there was significant non-sampling variation which could not be accounted

for by any observed characteristics of the studies; consequently, the sampling standard error was smaller than the total standard error. Comparisons of standard errors computed using the refined methods described above and those computed under the form of meta-analysis used in the previously published analysis showed that the latter generally underestimated the standard error of the average effect size estimate. This bias probably exaggerated the statistical significance of the differences among effect sizes in the earlier study.

Table 2 shows a breakdown of student performance across the science curricula studied in terms of criterion clusters (achievement, perceptions, etc.) and a composite measure (all criterion clusters combined). The mean effect sizes reported in the 1983 analysis were unweighted and treated multiple effect sizes from a single study as independent observations. In the re-analysis study effect sizes were averaged within studies or subgroups before aggregation; hence, the 136 effect sizes in the composite performance analysis represent independent subgroups. In additio the sampling standard error (SSE) and total standard error (TSE) of the mean effect size are reported for the re-analysis from which a confidence interval can be established.

Similar to the earlier results, the re-analysis shows that students in the new science curricula outperformed their traditional program counterparts on the composite level and three of the six criterion clusters. It is interesting to note that the mean effect sizes in the revised analysis vary from the original values in both directions. In one case (related measures), the mean actually changes direction. This pattern of changes in both directions was similar in all aggregations. The important conclusion to be drawn from Table 2 is that the new science curricula had a generally positive impact on student performance across criterion measures. Of particular significance is the mean for the

criterion of academic achievement. Opponents often criticized the new curricula for compromising content to obtain process skill and attitude gains. Table 2 clearly shows that such a compromise did not occur overall.

Table 3 provides a detailed analysis of student performance by grade level. With these data a clearer picture of the differentiated impact of the new programs begins to emerge. For example, elementary level programs had a significant impact only on student perceptions (0.23*). However, additional analyses of primary grade (K-3) and intermediate grade (4-6) performance data revealed statistically significant effects on achievement (1.39*) and process skills (0.41*) for K-3 students and perceptions (0.25*) for 4-6 students. At the junior high and high school levels the impact is extended to include achievement. Except for the apparent negative impact on the related skills component in the high school grades, we suspect that the architects of the new curricula would probably be very pleased with this pattern of performance -- students developed their process skills and interest in science at the elementary grade level and then increased their achievement and continued their process skill development in later grades.

The effect of the refined statistical procedures on the analysis are readily seen in Table 3. In several instances results are dramatically different (e.g., 4-6 composition and process skills, 7-9 perceptions and analytic skills, and 10-12 perceptions).

Factors contributing to observed differences in male and female interest and performance in science continue to be an important topic for discussion in curriuclum development. Table 4 contains a breakdown of student performance by class composition (nominally male classes are those in which more than 75% of the students are male; nominally female classes are those in which more than

75% of the students are female). The re-analysis shows that new science curricula had a significant positive effect on males but not on females at the composite performance level, even though the estimated effect size was larger for the females (0.56) than for the males (0.24). At the criterion cluster level, the data again show that male achievement and perceptions in new science curricula were significantly positive but the female performance on corresponding criteria were not. It is encouraging to note, however, that the analytic skills of the females improved significantly in the new science programs.

Here again, the impact of the refined statistical methods on the mean effect sizes is significant. While the earlier analysis showed that the new curricula had a nonsignificant impact on male perceptions and female analytic skills and a significant impact on male process skills and female perceptions, the re-analysis shows different effects in all these areas.

Table 5 contains student performance data broken down by school type. Of particular interest in this table are the data for urban schools. The re-analysis confirms that the new curricula apparently had a greater positive impact on students in urban settings than in either the suburban or rural settings on both the composite performance level and the achievement cluster. In addition, the urban student analytic skills data appear to be more positive than the subruban student data. Considering the urban school student densities and the myriad problems associated with inner-city teaching, it would seem foolish not to explore the components of these "new" science curricula which produced such positive gains.

Table 6 contains a break-down of results of student performance by student socio-economic status. The reanalysis reveals that the effects of the new science curricula on low SES students were not as positive as the unweighted

1983 analysis showed. None of the low SES results is significantly different from 0. Similar non-significant results appear in the results of the high SES students.

The results of the re-analysis on data for the mid-SES students generally confirm the 1983 results with the exception of the analytic skills cluster and the related skills cluster. The significantly positive results of the 1983 study (analytic skills = 0.23* and related skills = 0.24*) appear as non-significant effects in the reanalysis.

Table 7 contains results of the student performance by high school subject analysis. Here the pattern of results is not as clear. The new biology and physics curriucla appear to have been effective overall and especially effective in achievement and process skills, respectively. But the chemistry curricula produced no significantly positive effects. In fact, both the biology and chemistry curricula had a negative impact on student performance in the related skills area. While the strong showings of the biology curricula on achievement and the physics curricula on process skill development are encouraging and worthy of closer examination as revisions in thes subject areas are pursued, the weak showing of the new chemistry curricula suggest the approaches of the new programs were not particulary successful. Possible reasons for the poor showing may be that the "new" chemistry curricula were not actually very different from the traditional curricula of the period and/or that teachers were simply not prepared to implement the new curricula. It may be that a radically different approach to high school chemistry is needed to raise student performance and interest in this area. One such alternative currently receiving attention is the use of technology as a context for teaching chemistry and other traditional

school sciences. However, the data on the efficacy of a "science/technology/society" (STS) approach are inconclusive at this point.

Tables 8-12 contain results of student performance on criterion clusters by specific science curriculum. A summary of the results in terms of their statistical significance is presented in Table 13. Two of the curriculum projects (SCIS and CBA) produced significant positive effects in three of the performance areas while four of them (ESS, ESTPSI, IPS, MSP) produced significant positive effects in two fo the performance areas. Significant negative effects were produced in two projects (BSCS-Y and CHEM STUDY) in the same performance area--related skills.

Perhaps the most controversial activity associated with the new science curricula of the sixties and seventies (and the one frequently debated today) is teacher inservice training. Were the NSF sponosred inservice workshops critical in the implementation of the new curricula? Table 7 contains a breakdown of the composite student performance data that divides the sample of studies into those that provided teacher inservice and those that did not. The re-analysis shows that students in new programs taught by teachers having received special inservice training on the use of the materials significantly outperformed students in traditional programs. Where teachers received no inservice training, the results though positive are not statistically significant. These results are in constrast to the earlier data which showed students in the "no-inservice" classrooms having the same significant positive gains as the students in the "inservice classrooms.

The difference in these two analyses are critical considering the debate currently being waged regarding the resurrection of inservice workshops. The

data in Table 14 indicate that student performance is significantly enhanced when special teacher inservice accompanies curriculum implementation.

The break down of student performance by length of treatment (period of time new curriculum was implemented before post-testing occurred) shown in Table 15 is particularly interesting. The three time intervals selected for study represent approximately half a school year, not quite a full school year, and at least a full school year. A definite pattern emerges when student performance data are examined across the three treatment periods. The positive effect on student performance decreases in both number and magnitude as the treatment period increases. When student performance was measured after 14 weeks of exposure to one of the new science programs, signficantly positive effects appeared in four of the five performance areas. After 28 weeks of exposure, only one performance area (achievement) appeared significantly positive while another area (related skills) actually reversed and appeared significantly negative.

Tables 16-27 contain a breakdown of student performance on standardized tests for the various student, school, and curriculum characteristics featured in Tables 2-12. Generally, effect sizes based on these tests were less positive than when data from all test-types were included. For example, whereas the standardized test data yielded only 13 significantly positive student performance effect sizes, when all test-type data were included, 25 significantly positive effect sizes appeared. Interestingly enough, howver, some break-downs that yielded significantly positive effect sizes when the standardized test data were analyzed separately (e.g., predominantly male groups process skills [Table 18] or S-APA process skills [Table 24]) yielded non-significant results when all test-type data were analyzed.

IMPLICATIONS FOR PROGRAM DEVELOPMENT

Results from this study suggest that the science programs developed during the sixties and seventies were generally effective in improving student performance on cognitive measures and raising attitudes about science. But these data alone are not likely to convince school officials and agencies such as the National Science Foundation to fund extensive curriculum devlopment efforts. The best that can be hoped is that some lessons will be learned from these data about the cumulative experience of curriculum development and implementation.

The importance of teacher inservice discussed above is an example of such a lesson. It is interesting to note that fewer than 33% of the studies analyzed bothered to report whether or not the teachers had received any training prior to or during the study period. We strongly suspect that no inservice training was involved in these studies. Moreover, based on survey data of teachers (Shymansky and Aldridge, 1982), it is probably safe to assume that the 33% figure is a reasonable estimate of the rate of inservicing nationwide that took place during the period of these new science curricula. Clearly, this is a problem for both the developers of a new program and the school district which adopts it that must be avoided in future efforts.

Another important lesson which can be drawn from the data relates to the issue of student performance in related areas. Science curriculum developers could help the cause of science in the schools by building in clear ties with reading and mathematics activities. For years reading and mathematics educators have seen the potential of science as a context for teaching these skills. Science curriculum developers can captialize on this interest by building in direct reading and applied math activities. Future science programs cannot afford to be associated with a decline in reading and math as the data for grades K-6 and

10-12 in Table 3 show. The reality is that if it comes down to a question of progress in science at the expense of reading or math, science will lose every time.

CONCLUDING REMARKS

This re-analysis of research on the effectiveness of new science curricula is, in a sense, a replication of the 1983 meta-analysis in that it asks the same questions of the same population of studies as the earlier analysis. In that regard, results of the re-analysis generally support the conclusions of the earlier study: the new science curricula of the sixties and seventies were more effective than the traditional textbook programs of the time. At the detailed level, however, the refined methods used in the re-analysis reveal some critical differences from the earlier analysis. In the refined analysis only four criterion cluster mean effects were significantly positive (compared to all seven clusters in the earlier analysis), and then by a smaller margin. In one case, (related skills), the mean effect size changed from a 0.25 (significant at .05 level) to a -0.10. Similarly, throughout the various subgroup analyses performed, other changes in average effect sizes and their precision resulted.

REFERENCES

Anderson, R.T., Kahl, S.R., Glass, G.V., and Smith, M.L. "Science education: A meta-analysis of major questions." Journal of Research in Science Teaching, 1983, 20, 379-385.

Box, G.E.P. "Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variance and of correlation errors in the two way classification." Annals of Mathematical Statistics, 1954, 25, 484-498.

Glass, G.V., Peckham, P.D., and Sanders, J.R. "Consequences of failure to meet assumptions underlying fixed effects analysis of variance and covariance." Review of Educational Researcher, 1972, 42, 237-288.

Glass, G.V. "Primary, secondary, and meta-analysis of research." Educational Researcher, 1976, 5, 3-8.

Glass, G.V., McGaw, B., and Smith, M.L. Meta-analysis in social research. Beverly Hills: Sage Publications, 1981.

Hedges, L.V. "Distribution theory for Glass's estimator of effect size and related estimates." Journal of Educational Statistics, 1981, 6, 107-128.

Hedges, L.V. "Estimation of effect size from a series of independent experiments." Psychological Bulletin, 1982, 92, 490-499.

Hedges, L.V. "Fitting categorical models to effect sizes from a series of experiments." Journal of Educational Statistics, 1982, 7, 119-132.

Hedges, L.V. "Fitting continuous models to effect size data." Journal of Educational Statistics, 1982, 7, 245-270.

Hedges, L.V. and Olkin, I. Statistical methods for meta-analysis. New York: Academic Press, 1985.

Kraemer, H.C. "Theory of estimation and testing of effect sizes: Use in meta-analysis." Journal of Educational Statistics, 1983, 8, 93-101.

Madow, W.G. "One source of downward bias in the analysis of variance and covariance." Annals of Mathematical Statistics, 1948, 19, 351-359.

National Science Foundation. Program solicitation: Programs for elementary school science instruction. Washington, D.C.: NSF 86-4, 1986.

Presby, S. "Overly broad categories obscure important differences." American Psychologist, 1978, 33, 514-515.

Rosenthal, R. and Rubin, D.B. "A simple, general purpose display of magnitude of experimental effect." Journal of Educational Psychology, 1982, 74, 166-169.

Scheffe', H. The analysis of variance. New York: John Wiley, 1959.

Shymansky, J.A. and Aldridge, B. G. "The teacher crisis in secondary school science and mathematics." Educational Leadership, 1982, 40, 63-67.

Shymansky, J.A., Kyle, W.C., and Alport, J.M. "The effects of new science curricula on student performance." Journal of Research in Science Teaching, 1983, 20, 387-404.

TABLE 2

STUDENT PERFORMANCE ACROSS ALL CURRICULA BY CRITERION CLUSTER

CRITERION	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
COMPOSITE	340	0.34*	0.65	320	0.27*	0.03	136	0.26*	0.01	0.05
ACHIEVEMENT	130	0.37*	0.73	135	0.26*	0.04	84	0.30*	0.01	0.07
PERCEPTIONS	51	0.37*	0.49	36	0.26*	0.06	18	0.19*	0.03	0.06
PROCESS	56	0.39*	0.56	75	0.40*	0.06	47	0.33*	0.02	0.10
ANALYTIC	35	0.25*	0.44	27	0.19*	0.08	25	0.13	0.02	0.11
RELATED	48	0.25*	0.69	37	0.10	0.08	17	-0.10	0.03	0.15
OTHER	21	0.33*	0.51	10	0.28	0.20	9	0.10	0.06	0.30

*Significantly different from 0 at $p<0.05$.

TABLE 3

STUDENT PERFORMANCE BY CRITERION CLUSTER AND GRADE LEVEL

	UNWEIGHTED RESULTS (105 studies)			UNWEIGHTED RESULTS (81 studies)			REFINED RESULTS (81 studies)			
LEVEL	N	MEAN ES	SD	N	MEAN ES	TSE	N	MEAN ES	SSE	TSE
ELEMENTARY (K-6)										
COMPOSITE	124	0.31*	0.52	97	0.32*	0.06	44	0.23*	0.03	0.11
ACHIEVEMENT	27	0.37*	0.74	27	0.28*	0.11	20	0.21	0.04	0.17
PERCEPTIONS	29	0.28*	0.46	15	0.36*	0.08	8	0.23*	0.06	0.06
PROCESS	16	0.56*	0.59	23	0.62*	0.14	14	0.53	0.04	0.27
ANALYTIC	1	0.06	0.00	2	0.19	0.18	2	0.19	0.08	0.18
RELATED	37	0.17*	0.27	21	0.05	0.09	8	-0.13	0.07	0.15
JUNIOR HIGH (7-9)										
COMPOSITE	72	0.31*	0.62	59	0.36*	0.05	22	0.33*	0.03	0.09
ACHIEVEMENT	13	0.23*	0.34	19	0.36*	0.09	10	0.39*	0.04	0.13
PERCEPTIONS	11	0.59*	0.31	6	0.44*	0.14	3	0.33	0.07	0.22
PROCESS	18	0.23*	0.39	18	0.43*	0.06	12	0.39*	0.04	0.07
ANALYTIC	14	0.02	0.23	5	0.19	0.10	5	0.23*	0.04	0.09
RELATED	9	0.68	1.46	11	0.30	0.17	6	0.10	0.08	0.17
HIGH SCHOOL (10-12)										
COMPOSITE	132	0.38*	0.73	151	0.21*	0.04	64	0.25*	0.01	0.06
ACHIEVEMENT	83	0.37*	0.80	86	0.24*	0.06	51	0.30*	0.02	0.07
PERCEPTIONS	9	0.44	0.70	11	0.09	0.11	5	0.11*	0.04	0.04
PROCESS	19	0.43*	0.68	29	0.26*	0.07	18	0.22*	0.03	0.08
ANALYTIC	19	0.42*	0.50	19	0.19	0.11	17	0.08	0.02	0.13
RELATED	2	-0.23	0.38	5	-0.13*	0.05	3	-0.15*	0.04	0.05

*Significantly different from 0 at $p<0.05$.

TABLE 4

STUDENT PERFORMANCE BY CRITERION CLUSTER AND STUDENT GENDER

GENDER	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
MALE SAMPLE										
COMPOSITE	123	0.22*	0.45	125	0.20*	0.04	59	0.24*	0.02	0.06
ACHIEVEMENT	58	0.25*	0.49	64	0.19*	0.05	39	0.28*	0.02	0.07
PERCEPTIONS	12	-0.02	0.56	10	0.28	0.14	4	0.19*	0.06	0.06
PROCESS	13	0.16	0.36	17	0.29*	0.09	12	0.27	0.04	0.14
ANALYTIC	21	0.30	0.40	17	0.26*	0.11	15	0.19	0.03	0.15
RELATED	8	0.01	0.14	13	-0.04	0.04	6	-0.02	0.05	0.05
MIXED SAMPLE										
COMPOSITE	199	0.43*	0.71	183	0.31*	0.04	70	0.26*	0.02	0.08
ACHIEVEMENT	68	0.45*	0.88	66	0.33*	0.07	40	0.29*	0.02	0.10
PERCEPTIONS	34	0.51*	0.38	23	0.21*	0.07	13	0.18*	0.03	0.07
PROCESS	33	0.52*	0.64	55	0.45*	0.07	32	0.36*	0.03	0.13
ANALYTIC	9	0.31	0.56	9	0.04	0.09	9	0.06	0.03	0.11
RELATED	40	0.30*	0.74	24	0.17	0.11	11	-0.14	0.04	0.27
FEMALE SAMPLE										
COMPOSITE	19	0.25*	0.50	12	0.36	0.19	7	0.57	0.07	0.41
ACHIEVEMENT	4	0.55	0.88	5	0.31	0.43	5	0.70	0.08	0.73
PERCEPTIONS	5	0.32	0.45	3	0.54*	0.13	1	0.54	0.28	0.28
PROCESS	5	0.29*	0.23	3	0.28	0.30	3	0.32	0.10	0.38
ANALYTIC	5	-0.10	0.24	1	0.35	--	1	0.35*	0.11	0.11

*Significantly different from 0 at $p<0.05$.

TABLE 5

STUDENT PERFORMANCE BY CRITERION CLUSTER AND SCHOOL TYPE

TYPE	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
<u>RURAL</u>										
COMPOSITE	25	0.20*	0.44	25	0.36*	0.09	12	0.27	0.04	0.14
ACHIEVEMENT	9	0.34*	0.25	13	0.30	0.15	11	0.21	0.05	0.19
PERCEPTIONS	9	-0.07	0.58	1	0.48	--	1	0.48*	0.19	0.19
PROCESS	6	0.45*	0.23	10	0.43*	0.12	8	0.35*	0.06	0.17
<u>SUBURBAN</u>										
COMPOSITE	168	0.38*	0.74	163	0.24*	0.04	70	0.22*	0.02	0.08
ACHIEVEMENT	72	0.41*	0.85	59	0.31*	0.06	40	0.25*	0.02	0.11
PERCEPTIONS	19	0.46*	0.40	22	0.21*	0.06	12	0.12*	0.03	0.04
PROCESS	13	0.50*	0.85	23	0.40*	0.11	14	0.30	0.03	0.21
ANALYTIC	17	0.27*	0.45	19	0.13	0.09	17	0.14	0.03	0.12
RELATED	34	0.30*	0.79	33	0.11	0.08	13	-0.09	0.04	0.19
<u>URBAN</u>										
COMPOSITE	32	0.34*	0.47	14	0.66*	0.08	8	0.67*	0.04	0.09
ACHIEVEMENT	4	0.81*	0.41	7	0.85*	0.09	5	0.82*	0.05	0.05
PERCEPTIONS	2	0.64*	0.06	--	--	--	--	--	--	--
PROCESS	12	0.24	0.44	3	0.29*	0.05	3	0.31*	0.06	0.06
ANALYTIC	11	0.17	0.47	2	0.39*	0.05	2	0.40*	0.07	0.07
RELATED	2	0.41	0.47	--	--	--	--	--	--	--

*Significantly different from 0 at p<0.05.

TABLE 6

STUDENT PERFORMANCE BY CRITERION CLUSTER AND SOCIO-ECONOMIC-STATUS

STATUS	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
LOW SES										
COMPOSITE	4	0.63*	0.41	3	0.56*	0.24	2	0.25	0.18	0.33
ACHIEVEMENT	1	1.08	0.00	2	0.78*	0.18	1	0.78	0.42	0.42
PROCESS	1	0.64	0.00	1	0.13	--	1	0.13	0.20	0.20
RELATED	2	0.41	1.08	--	--	--	--	--	--	--
MID-SES										
COMPOSITE	302	0.28*	0.51	291	0.28*	0.03	127	0.28*	0.01	0.05
ACHIEVEMENT	105	0.27*	0.49	124	0.25*	0.04	77	0.32*	0.01	0.07
PERCEPTIONS	49	0.32*	0.44	32	0.29*	0.06	16	0.20*	0.03	0.06
PROCESS	49	0.33*	0.46	68	0.43*	0.06	42	0.36*	0.02	0.11
ANALYTIC	33	0.23*	0.42	23	0.24*	0.09	21	0.14	0.02	0.11
RELATED	46	0.24*	0.70	35	0.11	0.08	16	-0.08	0.04	0.17
HIGH SES										
COMPOSITE	19	0.99*	1.34	26	0.11	0.07	7	-0.04	0.05	0.14
ACHIEVEMENT	11	1.00*	1.59	9	0.30	0.17	6	-0.05	0.05	0.27
PERCEPTIONS	2	1.40	0.49	4	-0.02	0.03	2	-0.01	0.13	0.13
PROCESS	4	1.00	1.31	6	0.19	0.11	4	0.04	0.06	0.19
ANALYTIC	2	0.50	0.82	4	-0.09	0.09	4	-0.10	0.08	0.09

*Significantly different from 0 at $p<0.05$.

TABLE 7

STUDENT PERFORMANCE BY CRITERION CLUSTER AND SCIENCE SUBJECT AREA

SUBJECT	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
EARTH SCIENCE										
COMPOSITE	20	0.13*	0.29	7	0.04	0.10	2	0.07	0.06	0.10
ACHIEVEMENT	4	-0.07	0.32	4	-0.04	0.17	1	-0.04	0.07	0.07
PERCEPTIONS	1	0.11	[illegible]	1	0.11	--	1	0.11	0.07	0.07
PROCESS	8	0.22	0.39	1	0.21	--	1	0.21	0.12	0.12
ANALYTIC	7	0.16*	0.18	1	0.15	--	1	0.15*	0.07	0.07
PHYSICAL SCIENCE										
COMPOSITE	34	0.18*	0.30	38	0.26*	0.05	12	0.33*	0.03	0.07
ACHIEVEMENT	9	0.31*	0.47	10	0.45*	0.11	7	0.47*	0.03	0.17
PERCEPTIONS	8	0.31	0.16	10	0.09	0.06	6	0.11*	0.05	0.05
PROCESS	10	0.08	0.34	14	0.26*	0.08	9	0.34*	0.03	0.09
ANALYTIC	7	-0.10	0.21	4	0.23*	0.11	4	0.26	0.04	0.13
BIOLOGY										
COMPOSITE	47	0.60*	0.91	54	0.30*	0.06	29	0.33*	0.02	0.09
ACHIEVEMENT	29	0.59	1.04	31	0.40*	0.08	24	0.43*	0.02	0.10
PERCEPTIONS	4	0.82	0.72	6	0.04	0.04	3	0.11	0.06	0.06
PROCESS	6	0.90	0.96	9	0.24	0.13	7	0.22	0.07	0.19
ANALYTIC	7	0.46	0.58	7	0.24	0.24	7	-0.05	0.04	0.33

(continued on next page)

TABLE 7 (continued)

<u>CHEMISTRY</u>										
COMPOSITE	49	0.14*	0.38	56	0.11	0.06	19	0.10	0.03	0.10
ACHIEVEMENT	33	0.16*	0.40	31	0.04	0.07	17	0.13	0.03	0.11
PERCEPTIONS	4	0.15	0.69	4	0.15	0.33	1	0.15	0.29	0.29
PROCESS	6	0.02	0.25	12	0.22	0.13	6	0.13	0.05	0.30
ANALYTIC	6	0.28	0.29	7	0.28	0.19	7	0.26	0.04	0.26
<u>PHYSICS</u>										
COMPOSITE	33	[illegible]	0.66	37	0.27*	0.09	18	0.28*	0.03	0.12
ACHIEVEMENT	23	0.50*	0.77	23	0.35*	0.14	12	0.35	0.03	0.18
PROCESS	7	0.33	0.40	6	0.33*	0.09	5	0.31*	0.05	0.11
ANALYTIC	6	0.53	0.61	5	-0.01	0.07	3	-0.03	0.05	0.05

*Significantly different from 0 at $p<0.05$.

TABLE 8

STUDENT PERFORMANCE ON ACHIEVEMENT TESTS BY CURRICULUM

CURRICULUM	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
ESS	3	0.09	0.12	4	0.03	0.17	4	0.04	0.08	0.18
SCIS	5	1.00	0.91	3	0.89*	0.15	2	1.09*	0.13	0.13
S-APA	12	0.17	0.58	13	0.06	0.14	11	0.03	0.04	0.22
MINNEMAST	2	1.51	1.35	1	1.72	--	1	1.72*	0.16	0.16
[illegible]TPSI	3	0.28	0.27	12	0.24*	0.08	4	0.26*	0.08	0.08
FHESP	1	-0.06	0.00	5	0.47	0.27	3	0.24	0.08	0.32
IPS	3	0.03	0.26	3	0.23	0.16	2	0.28	0.10	0.24
ESCP	6	0.19	0.49	6	0.21	0.19	2	0.17	0.06	0.42
IME	2	-0.11	0.30	2	0.19*	0.04	1	0.19	0.12	0.12
MSP	1	0.42	0.00	1	0.49	--	1	0.49*	0.05	0.05
BSCS-S	2	0.02	0.01	4	0.41	0.26	2	0.44	0.12	0.27
BSCS-Y	19	0.45*	0.54	19	0.43*	0.09	18	0.47*	0.02	0.11
BSCS-B	2	3.94*	0.33	2	1.01*	0.01	1	1.01*	0.26	0.26
BSCS-G	2	0.17	0.24	2	0.01	0.01	1	0.01	0.08	0.08
BSCS-A	4	0.09	0.28	4	0.08	0.14	2	0.11	0.12	0.12
CHEM STUDY	23	0.12	0.40	24	-0.10	0.06	13	0.03	0.03	0.10
CBA	10	0.27	0.41	7	0.51*	0.18	4	0.53*	0.06	0.10
PSSC	23	0.51*	0.77	23	0.35*	0.14	12	0.34	0.03	0.18

*Significantly different from 0 at $p<0.05$.

TABLE 9

STUDENT PERFORMANCE ON PERCEPTIONS TESTS BY CURRICULUM

CURRICULUM	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
ESS	1	0.51	0.00	1	0.47*	--	1	0.47*	0.19	0.19
SCIS	14	0.08	0.52	10	0.37*	0.10	4	0.24*	0.09	0.09
S-APA	6	0.39*	0.28	4	0.30*	0.17	3	0.18	0.07	0.18
HSP	4	0.66*	0.18	4	0.63	0.08	1	0.63*	0.10	0.10
IPS	2	0.23*	0.02	1	-0.08	--	1	-0.08	0.14	0.14
ESCP	1	0.11	0.00	1	0.11	--	1	0.11	0.07	0.07
IME	2	0.55	0.14	1	0.20	--	1	0.20	0.12	0.12
BSCS-Y	1	1.05	0.00	2	0.02	0.05	1	0.02	0.16	0.16
BSCS-B	2	0.25	0.29	2	-0.05	0.00	1	-0.05	0.20	0.20
BSCS-G	1	1.75	0.00	2	0.14	0.01	1	0.14*	0.06	0.06
CBA	4	0.16	0.69	4	0.15	0.33	1	0.15	0.28	0.28
PSNS	1	0.15	0.00	4	0.07	0.08	2	0.15	0.07	0.14

*Significantly different from 0 at p<0.05.

TABLE 10

STUDENT PERFORMANCE ON PROCESS SKILL TESTS BY CURRICULUM

CURRICULUM	UNWEIGHTED RESULTS (105 studies)			UNWEIGHTED RESULTS (81 studies)			REFINED RESULTS (81 studies)			
	N	MEAN ES	SD	N	MEAN ES	TSE	N	MEAN ES	SSE	TSE
ESS	4	0.47*	0.18	12	0.43*	0.10	4	0.43*	0.12	0.12
SCIS	6	0.56*	0.36	10	0.64*	0.21	6	0.42	0.06	0.48
S-APA	3	1.08	1.28	3	1.30*	0.60	2	0.89	0.08	1.11
USMES	1	0.29	0.00	1	0.25	--	1	0.25	0.32	0.32
ESTPSI	3	0.50	0.30	4	0.64*	0.13	4	0.63*	0.08	0.13
FHESP	--	--	--	5	0.54*	0.10	3	0.39*	0.08	0.16
ISCS	1	0.30	0.00	3	0.09	0.11	1	0.09	0.19	0.19
IPS	5	-0.08	0.30	4	0.26*	0.04	4	0.29*	0.05	0.05
ESCP	8	0.22	0.39	1	0.20	--	1	0.20	0.12	0.12
IME	--	--	--	1	0.66	--	1	0.66*	0.13	0.13
BSCS-Y	4	0.72	0.71	7	0.22	0.16	6	0.21	0.07	0.23
BSCS-B	1	2.45	0.00	2	0.26*	0.02	1	0.26	0.21	0.21
CHEM STUDY	5	-0.03	0.22	10	0.21	0.15	5	0.08	0.05	0.40
CBA	1	0.34	0.00	2	0.26	0.45	1	0.26*	0.09	0.09
PSSC	5	0.35	0.49	4	0.37*	0.12	4	0.33*	0.05	0.13
HPP	2	0.28*	0.02	2	0.24*	0.07	1	0.24*	0.09	0.09
PSNS	2	0.09	0.01	4	0.02	0.05	2	0.05	0.07	0.07

*Significantly different from 0 at $p<0.05$.

TABLE 11

STUDENT PERFORMANCE ON ANALYTIC SKILL TEST BY CURRICULUM

	UNWEIGHTED RESULTS (105 studies)			UNWEIGHTED RESULTS (81 studies)			REFINED RESULTS (81 studies)			
RRICULUM	N	MEAN ES	SD	N	MEAN ES	TSE	N	MEAN ES	SSE	TSE
;	--	--	--	1	0.01	--	1	0.01	0.10	0.10
.PA	1	0.06	0.00	1	0.36	--	1	0.36*	0.10	0.10
S	1	0.07	0.00	1	-0.06	--	1	-0.06	0.19	0.19
	5	-0.15	0.22	2	0.39*	0.05	2	0.40*	0.07	0.07
:P	7	0.16*	0.18	1	0.15	--	1	0.15*	0.07	0.07
'ES	1	0.01	0.00	1	0.03	--	1	0.03	0.11	0.11
)	1	0.12	0.00	1	0.19	--	1	0.19*	0.05	0.05
:S-S	1	0.29	0.00	1	0.23	--	1	0.23*	0.09	0.09
:S-Y	5	0.42	0.57	3	-0.00	0.33	3	-0.28	0.05	0.43
:S-B	2	0.94	0.70	2	0.69	0.73	2	0.74	0.14	0.74
:S-G	1	-0.18	0.00	1	0.02	--	1	0.02	0.13	0.13
:M STUDY	5	0.30	0.32	6	0.24	0.21	6	0.19	0.04	0.32
\	1	0.21	0.00	1	0.53	--	1	0.53*	0.09	0.09
C	6	0.53	0.61	5	-0.00	0.07	3	-0.02	0.05	0.05

}nificantly different from 0 at p<0.05.

TABLE 12

STUDENT PERFORMANCE ON RELATED SKILL TESTS BY CURRICULUM

CURRICULUM	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
SCIS	13	0.21*	0.15	7	0.16*	0.05	1	0.16	0.19	0.19
S-APA	18	0.10	0.29	12	-0.17*	0.06	4	-0.18	0.07	0.09
USMES	12	0.66	1.25	10	0.51*	0.22	6	0.32	0.13	0.26
ISCS	2	-0.03	0.09	2	-0.04	0.07	2	-0.03	0.15	0.15
MSP	1	0.11	0.00	1	0.11	--	1	0.11	0.16	0.16
BSCS-Y	1	-0.50	0.00	1	-0.17	--	1	-0.17*	0.06	0.06
CHEM STUDY	--	--	--	2	-0.21*	0.04	1	-0.21*	0.08	0.08
PSSC	1	0.04	0.00	2	-0.02	0.03	1	-0.02	0.09	0.09

*Significantly different from 0 at $p<0.05$.

TABLE 13

SUMMARY OF STUDENT PERFORMANCE BY CURRICULUM IN TERMS OF SIGNIFICANT EFFECT SIZES YIELDED IN REANALYSES

CURRICULUM	ACHIEVEMENT	PERCEPTIONS	PROCESS	ANALYTIC	RELATED
ESS	0	+	+	0	?
SCIS	+	+	0	+	0
S-APA	0	0	0	?	0
USMES	?	?	0	?	?
MINNEMAST	+	?	?	?	0
ESTPSI	+	?	+	?	?
FHESP	0	?	+	?	?
HSP	?	+	?	?	?
IPS	0	0	+	+	?
ISCS	?	?	0	0	0
ESCP	0	0	0	+	?
CE/ES	?	?	?	0	?
IME	0	0	+	?	?
MSP	+	?	?	+	0
BSCS-S	0	?	?	+	0
BSCS-Y	+	0	0	0	-
BSCS-B	+	0	0	0	?
BSCS-G	0	+	?	0	?
BSCS-A	0	?	?	?	?
CHEM STUDY	0	?	0	0	-
CBA	+	0	+	+	?
PSSC	0	?	+	0	0
HPP	?	?	+	?	?
PSNS	?	0	0	?	?

+ denotes significant positive effect
- denotes significant negative effect
0 denotes nonsignificant effect
? denotes no data available

TABLE 14

STUDENT PERFORMANCE BY CRITERION CLUSTER AND LEVEL OF TEACHER INSERVICE

LEVEL	UNWEIGHTED RESULTS (105 studies) N	MEAN ES	SD	UNWEIGHTED RESULTS (81 studies) N	MEAN ES	TSE	REFINED RESULTS (81 studies) N	MEAN ES	SSE	TSE
INSERVICE										
COMPOSITE	112	0.23*	0.45	117	0.25*	0.04	49	0.27*	0.02	0.07
ACHIEVEMENT	40	0.22*	0.41	55	0.17*	0.06	33	0.28*	0.03	0.10
PERCEPTIONS	19	0.16	0.48	7	0.29*	0.14	6	0.23*	0.07	0.10
PROCESS	27	0.32*	0.55	32	0.47*	0.09	21	0.36*	0.03	0.14
ANALYTIC	15	0.07	0.22	9	0.13	0.07	7	0.16	0.03	0.09
RELATED	5	0.12	0.30	4	0.50	0.33	4	0.15	0.07	0.46
NO INSERVICE										
COMPOSITE	14	0.50*	0.32	9	0.24*	0.11	5	0.24	0.05	0.15
ACHIEVEMENT	9	0.46*	0.39	7	0.23	0.15	4	0.21	0.06	0.19
PROCESS	1	0.32	0.00	1	0.33	--	1	0.33*	0.08	0.08
ANALYTIC	2	0.62	0.18	1	0.22	--	1	0.22	0.14	0.14

*Significantly different from 0 at $p<0.05$.

TABLE 15

STUDENT PERFORMANCE BY CRITERION CLUSTER AND LENGTH OF TREATMENT

LENGTH	N	MEAN ES	SSE	TSE
<u>LESS THAN 14 WEEKS</u>				
COMPOSITE	15	0.41*	0.04	0.18
ACHIEVEMENT	8	0.62*	0.06	0.23
PERCEPTIONS	2	0.44*	0.16	0.16
PROCESS	1	0.53*	0.18	0.18
ANALYTIC	3	0.17*	0.06	0.06
RELATED	10	0.07	0.06	0.23
<u>BETWEEN 14 AND 25 WEEKS</u>				
COMPOSITE	69	0.23*	0.02	0.05
ACHIEVEMENT	37	0.25*	0.02	0.07
PERCEPTIONS	5	0.13*	0.05	0.05
PROCESS	25	0.36*	0.03	0.08
ANALYTIC	16	0.23	0.03	0.15
<u>GREATER THAN 28 WEEKS</u>				
COMPOSITE	52	0.26*	0.01	0.10
ACHIEVEMENT	39	0.29*	0.02	0.11
PERCEPTIONS	11	0.20	0.04	0.10
PROCESS	21	0.29	0.03	0.21
ANALYTIC	6	-0.06	0.03	0.11
RELATED	7	-0.17*	0.04	0.07

*Significantly different from 0 at $p<0.05$.

TABLE 16

STUDENT PERFORMANCE ON STANDARDIZED TESTS ACROSS ALL CURRICULA

BY CRITERION CLUSTER

CRITERION	N	MEAN ES	SSE	TSE
COMPOSITE	88	0.22*	0.01	0.05
ACHIEVEMENT	54	0.24*	0.02	0.08
PERCEPTIONS	13	0.22*	0.04	0.08
PROCESS	30	0.31*	0.02	0.04
ANALYTIC	22	0.14	0.00	0.11
RELATED	10	-0.11*	0.04	0.05

*Significantly different from 0 at $p<0.05$.

TABLE 17

STUDENT PERFORMANCE ON STANDARDIZED TESTS BY CRITERION CLUSTER AND GRADE LEVEL

LEVEL	N	MEAN ES	SSE	TSE
ELEMENTARY (K-6)				
COMPOSITE	24	0.11	0.03	0.11
ACHIEVEMENT	15	0.04	0.04	0.16
PERCEPTIONS	7	0.25*	0.06	0.06
PROCESS	4	0.50*	0.07	0.18
ANALYTIC	2	0.19	0.08	0.18
RELATED	4	-0.17	0.00	0.13
JUNIOR HIGH (7-9)				
COMPOSITE	18	0.32*	0.03	0.08
ACHIEVEMENT	8	0.40*	0.04	0.17
PERCEPTIONS	3	0.33	0.07	0.21
PROCESS	10	0.37*	0.04	0.07
ANALYTIC	5	0.23*	0.04	0.11
RELATED	4	0.00	0.08	0.08
HIGH SCHOOL (10-12)				
COMPOSITE	43	0.19*	0.02	0.08
ACHIEVEMENT	29	0.24*	0.02	0.11
PERCEPTIONS	3	0.11	0.06	0.06
PROCESS	15	0.21*	0.03	0.07
ANALYTIC	14	0.10	0.03	0.18
RELATED	2	-0.13	0.05	0.09

*Significantly different from 0 at $p<0.05$.

TABLE 18

STUDENT PERFORMANCE ON STANDARIZED TESTS BY CRITERION CLUSTER

AND STUDENT GENDER

GENDER	N	MEAN ES	SSE	TSE
MALE SAMPLE				
COMPOSITE	42	0.22*	0.02	0.07
ACHIEVEMENT	25	0.26*	0.03	0.09
PERCEPTIONS	3	0.19*	0.06	0.06
PROCESS	8	0.27*	0.05	0.08
ANALYTIC	12	0.22	0.00	0.17
RELATED	6	-0.02	0.05	0.05
MIXED SAMPLE				
COMPOSITE	40	0.21*	0.02	0.06
ACHIEVEMENT	25	0.23*	0.02	0.11
PERCEPTIONS	9	0.22*	0.05	0.10
PROCESS	19	0.33*	0.03	0.05
ANALYTIC	9	0.06	0.03	0.11
RELATED	4	-0.21	0.06	0.14
FEMALE SAMPLE				
COMPOSITE	6	0.31	0.09	0.50
ACHIEVEMENT	4	0.01	0.19	0.54
PERCEPTIONS	1	0.54	0.28	0.28
PROCESS	3	0.32	0.10	0.39
ANALYTIC	1	0.35*	0.10	0.10

*Significantly different from 0 at $p<0.05$.

TABLE 19

STUDENT PERFORMANCE ON STANDARDIZED TESTS BY CRITERION CLUSTER AND SCHOOL TYPE

TYPE	N	MEAN ES	SSE	TSE
<u>RURAL</u>				
COMPOSITE	10	0.37*	0.06	0.15
ACHIEVEMENT	10	0.29	0.06	0.19
PERCEPTIONS	1	0.48*	0.19	0.19
PROCESS	7	0.38	0.08	0.19
<u>SUBURBAN</u>				
COMPOSITE	46	0.18*	0.02	0.08
ACHIEVEMENT	27	0.19	0.03	0.14
PERCEPTIONS	10	0.13*	0.04	0.06
PROCESS	10	0.32*	0.04	0.10
ANALYTIC	14	0.15	0.03	0.14
RELATED	7	-0.13	0.04	0.07
<u>URBAN</u>				
COMPOSITE	4	0.55*	0.06	0.17
ACHIEVEMENT	2	0.88*	0.09	0.24
PROCESS	3	0.31*	0.06	0.06
ANALYTIC	2	0.40*	0.07	0.07

*Significantly different from 0 at $p<0.05$.

TABLE 20

STUDENT PERFORMANCE ON STANDARDIZED TESTS BY CRITERION CLUSTER AND SOCIO-ECONOMIC-STATUS

STATUS	N	MEAN ES	SSE	TSE
<u>MID-SES</u>				
COMPOSITE	85	0.23*	0.01	0.05
ACHIEVEMENT	51	0.26*	0.02	0.08
PERCEPTIONS	11	0.24*	0.04	0.08
PROCESS	27	0.36*	0.03	0.04
ANALYTIC	21	0.14	0.02	0.11
RELATED	9	-0.07	0.05	0.05
<u>HIGH SES</u>				
COMPOSITE	3	-0.10	0.07	0.19
ACHIEVEMENT	3	-0.08	0.07	0.52
PERCEPTIONS	2	-0.01	0.13	0.13
PROCESS	3	-0.02	0.07	0.08
ANALYTIC	1	-0.16	0.17	0.17
RELATED	1	-0.21*	0.07	0.07

*Significantly different from 0 at $p<0.05$.

TABLE 21

STUDENT PERFORMANCE ON STANDARDIZED TESTS BY CRITERION CLUSTER AND SCIENCE SUBJECT AREA

SUBJECT	N	MEAN ES	SSE	TSE
EARTH SCIENCE				
COMPOSITE	2	0.17*	0.06	0.06
PROCESS	1	0.21	0.12	0.12
ANALYTIC	1	0.15*	0.07	0.07
PHYSICAL SCIENCE				
COMPOSITE	9	0.36*	0.03	0.07
ACHIEVEMENT	6	0.42*	0.04	0.12
PERCEPTIONS	3	0.08	0.06	0.08
PROCESS	7	0.43*	0.05	0.10
ANALYTIC	4	0.26	0.04	0.13
BIOLOGY				
COMPOSITE	20	0.18	0.03	0.13
ACHIEVEMENT	15	0.28	0.04	0.17
PERCEPTIONS	3	0.11	0.06	0.06
PROCESS	7	0.22	0.07	0.19
ANALYTIC	7	-0.05	0.04	0.32
CHEMISTRY				
COMPOSITE	10	0.15	0.04	0.10
ACHIEVEMENT	8	0.10	0.04	0.11
PROCESS	4	0.07	0.05	0.05
ANALYTIC	4	0.35	0.05	0.31
RELATED	1	-0.21*	0.08	0.08

(continued on next page)

TABLE 21 (continued)

PHYSICS				
COMPOSITE	14	0.26	0.03	0.15
ACHIEVEMENT	8	0.33	0.04	0.26
PROCESS	4	0.39*	0.05	0.07
ANALYTIC	3	-0.03	0.05	0.05
RELATED	1	-0.02	0.09	0.09

*Significantly different from 0 at $p<0.05$.

TABLE 22

STUDENT PERFORMANCE ON STANDARDIZED TESTS OF ACHIEVEMENT BY CURRICULUM

CURRICULUM	N	MEAN ES	SSE	TSE
ESS	4	0.04	0.08	0.18
S-APA	11	0.03	0.05	0.23
FHESP	3	0.24	0.09	0.32
IPS	2	0.28	0.11	0.25
ESCP	1	0.74	0.12	0.12
IME	1	0.20	0.13	0.13
MSP	1	0.49*	0.06	0.06
BSCS-S	2	0.44	0.12	0.27
BSCS-Y	9	0.33	0.04	0.25
BSCS-B	1	1.01	0.27	0.27
BSCS-G	1	0.01	0.09	0.09
BSCS-A	2	0.11	0.13	0.13
CHEM STUDY	5	0.01	0.06	0.18
CBA	3	0.32*	0.08	0.08
PSSC	8	0.33	0.04	0.26

*Significantly different from 0 at $p<0.05$.

TABLE 23

STUDENT PERFORMANCE ON STANDARDIZED TESTS OF PERCEPTIONS

BY CURRICULUM

CURRICULUM	N	MEAN ES	SSE	TSE
ESS	1	0.48*	0.19	0.19
SCIS	3	0.35	0.14	0.14
S-APA	3	0.19	0.07	0.19
HSP	1	0.63*	0.11	0.10
IPS	1	-0.09	0.14	0.14
IME	1	0.20	0.13	0.13
BSCS-Y	1	0.02	0.17	0.17
BSCS-B	1	-0.06	0.21	0.21
BSCS-G	1	0.14*	0.06	0.06

*Significantly different from 0 at $p<0.05$.

TABLE 24

STUDENT PERFORMANCE ON STANDARDIZED TESTS OF PROCESS SKILLS

BY CURRICULUM

CURRICULUM	N	MEAN ES	SSE	TSE
ESS	4	0.43*	0.12	0.12
S-APA	1	0.72*	0.09	0.09
FHESP	3	0.39*	0.09	0.17
ISCS	1	0.09	0.20	0.20
IPS	4	0.29*	0.06	0.06
ESCP	1	0.21	0.12	0.12
IME	1	0.67*	0.13	0.13
BSCS-Y	6	0.21	0.08	0.24
BSCS-B	1	0.27	0.21	0.21
CHEM STUDY	3	-0.02	0.07	0.07
CBA	1	0.26*	0.09	0.09
PSSC	3	0.44*	0.06	0.06
HPP	1	0.24*	0.10	0.10

*Significantly different from 0 at $p<0.05$.

TABLE 25

STUDENT PERFORMANCE ON STANDARDIZED TESTS OF ANALYTIC SKILLS

BY CURRICULUM

CURRICULUM	N	MEAN ES	SSE	TSE
ESS	1	0.01	0.11	0.11
S-APA	1	0.37	0.11	0.11
ISCS	1	-0.07	0.20	0.20
IPS	2	-0.40	0.07	0.07
ESCP	1	0.15	0.07	0.07
CE/EE	1	0.03	0.12	0.12
MSP	1	0.20	0.06	0.06
BSCS-S	1	0.24	0.09	0.09
BSCS-Y	3	-0.28	0.05	0.44
BSCS-B	2	0.75	0.14	0.74
BSCS-G	1	0.03	0.14	0.14
CHEM STUDY	3	0.29	0.05	0.47
CBA	1	0.53	0.09	0.09
PSSC	3	-0.03	0.05	0.05

*Significantly different from 0 at $p<0.05$.

TABLE 26

STUDENT PERFORMANCE ON STANDARDIZED TESTS OF RELATED SKILLS

BY CURRICULUM

CURRICULUM	N	MEAN ES	SSE	TSE
SCIS	1	0.16	0.19	0.19
S-APA	4	-0.18	0.07	0.09
ISCS	2	-0.03	0.15	0.15
MSP	1	0.11	0.16	0.16
CHEM STUDY	1	-0.21	0.08	0.08
PSSC	1	-0.02	0.09	0.09

*Significantly different from 0 at $p<0.05$.

TABLE 27

STUDENT PERFORMANCE ON STANDARDIZED TESTS BY CRITERION CLUSTER

AND LEVEL OF TEACHER INSERVICE

LEVEL	N	MEAN ES	SSE	TSE
INSERVICE				
COMPOSITE	41	0.22*	0.02	0.07
ACHIEVEMENT	25	0.15	0.04	0.11
PERCEPTIONS	6	0.23*	0.07	0.11
PROCESS	19	0.37*	0.03	0.06
ANALYTIC	7	0.16	0.03	0.09
RELATED	3	-0.04	0.06	0.06
NO INSERVICE				
COMPOSITE	3	0.20	0.06	0.26
ACHIEVEMENT	3	0.21	0.06	0.26
ANALYTIC	1	0.22	0.14	0.14

*Significantly different from 0 at $p<0.05$.